

NESCent Anatomy Ontologies Course Notes

July 30 to August 3, 2012

Course Wiki https://academy.nescent.org/wiki/Anatomy_ontologies

Course Materials <http://code.google.com/p/nescent-anatomy-course/>

Materials, including presentations, tutorials, exercises, and ontology files, may be checked out from the Google Code SVN repository following the instructions on the Source tab. Location of relevant documents in the `/svn/trunk` directory is provided throughout the notes. Slides and tutorials in PDF format can also be downloaded at: <http://bit.ly/RuwZlz> (62 MB zip archive).

Participant

John Cork
Alex Dececchi
Robert Druzinsky
Eric Edsinger-Gonzales
Alexis Hazbun
Nizar Ibrahim
Micheal (Andrew) Jansen
Emmanuel Maxime
Laura Moore
Gaurav Vaidya
Christian Wirkner
Rambert Yan

Institution

LSU Health Sciences Center, New Orleans
University of South Dakota
University of Illinois, Chicago
University of California at Berkeley
Hollings Marine Lab., Coll. of Charleston
University of Chicago
Arizona State University
University of Louisiana at Lafayette
Oregon State University
University of Colorado at Boulder
Universitaet Rostock, Germany
New Jersey Institute of Technology

Remote participants

Yvonne Bradford
Terry Hayamizu
VG Ponferrada
ZFIN, University of Oregon
The Jackson Laboratory
Xenbase, Cincinnati Children's Hospital MC

Instructors

Matt Yoder (MY)
Melissa Haendel (MH)
Jim Balhoff (JB)
Carlo Torniai (CT)
Erik Segerdell
Illinois Natural History Survey
Oregon Health & Science University
NESCent
Oregon Health & Science University
Oregon Health & Science University



Phenotype RCN



THE NATIONAL CENTER FOR
BIOMEDICAL ONTOLOGY

Syllabus

(This is the original schedule. Times and activities were adjusted as needed during the course of the week.)

Breaks are not currently indicated, but we anticipate them to occur at natural stopping points.			
Hands on activities are indicated in green. Pairs activities are indicated in the right hand column.			
A student assignment to create a 5 min presentation on any ontological topic or problem of interest is due on the last day.			
Outputs of exercises will be due in SVN. The instructors will be available for consultation on an as needed basis.			
	Time	Topic/Activity	Pairs?
Monday			
	Read for Mon: http://www.pensoft.net/journals/ihr/article/2961/a-hymenopterists		
	9am-9:20am	NESCent Introduction	
	9:20am-10am	Introductions -- who are we? Goals of the course.	
	10am-11am	Overview. The lifespan of an anatomy ontology.	
	11am-12pm	An introduction to (anatomy) ontologies.	
	12pm-1pm	lunch	
hands-on	1pm-2pm	SVN and version tracking.	
	2pm-2:30pm	Ontological communities and shared principles, trackers, lists, ontology metadata, writing good requests.	
hands-on	2:30pm-5pm	Modelling, visualizing, and planning your ontology.	yes
Tuesday			
	Read for Tues: http://www.sciencedirect.com/science/article/pii/S0169534707001048		
	9am-9:30am	Anatomy of an OBO file, Intro to OBOEdit.	
hands-on	9:30am-10:30am	Navigating an ontology in OBO edit.	
	10:30am-12pm	Basic ontology creation in OBO edit.	yes
	12pm-1pm	lunch	
	1pm-2pm	Spatial Ontology Exercise.	yes
	2pm-3pm	Using anatomy ontologies for annotation. Linking in genetic data.	
hands-on	3pm-4pm	Curating evolutionary phenotype data using Phenex.	
	4pm-5pm	Student presentation and discussion of their modeling.	yes
Wednesday			
	9am-9:30am	Intro to OWL.	
hands-on	9:30am-12pm	The big Protégé tutorial.	
	12pm-1pm	lunch	
	1pm-3pm	Continues with prior tutorial based on time. Explore anatomy ontology exercises.	
hands-on	3pm-4:45pm	Work on your own ontology.	
	4:45pm-5pm	Discussion.	

Thursday			
	Read for Thurs: http://genomebiology.com/2012/13/1/R5/		
hands-on	9am-10am	OWL: imports and external ontologies.	
	10am-10:30am	Intro to Ontobee and imports.	
	10:30am-11pm	Ontology alignment, interoperability, reuse, building compositional anatomical entities from different ontologies.	
	11am-12pm	Mireoting, Ontofox.	
	12pm-1pm	lunch	
hands-on	1pm-2pm	MIREOTing, Ontofox.	
	2pm-3pm	1:1 hands on with your own ontology: Referencing external sources and ontologies.	
	3pm-3:30pm	Protégé plugins. Image annotation plugin. Phenotype annotations in Protege sensu HAO.	
hands-on	3:30pm-4pm	Adding images to your ontology. Annotating classes.	
	4pm-5pm	Spatial exercise.	
Friday			
	Read for Fri: http://www.plosbiology.org/article/info%3Adoi%2F10.1371%2Fjournal.pbio.1000247		
	9am-10am	Discuss group projects.	pairs modeling, OWL files, any projects welcome
hands-on	10am-12pm	Releasing ontologies in multiple formats using the Oort tool.	
	12pm-1pm	lunch	
	1pm-2pm	Homology 1, issues and navigating through ontologies that formalize homology hypotheses.	
	1:30pm-2pm	Homology 2, modern sounds in homology reasoning.	
hands-on	2pm-3:30pm	Student presentations.	
	3:30-4pm	Karen -- Course Evaluation.	
	4pm-5pm	NESCent-wide reception.	
		Dinner and Social.	

Software

(green) Required		(peach) Optional	(blue) Likely required, but installation may be tricky, so don't panic.
Title	OS	Source	Notes
SVN client			We recommend either a command line version or SmartSVN, though we are providing a tutorial with reference to the command line version.
Command line SVN (recommended for workshop)	windows	http://www.collab.net/downloads/subversion	
	mac	built in(?)	Try "svn info" from a command prompt, if you get "svn: '.' is not a working copy" then you're set up.
	linux	built in(?)	See above.
Smart SVN	all	http://www.syntevo.com/smartsvn/index.html	
Tortoise SVN	windows	http://tortoisesvn.net/	
OBO Edit v. 2.2	all	http://sourceforge.net/projects/geneontology/files/OBO-Edit 2 %5BLatest versions%5D/	
Text editor (student choice, *NOT* MS Word): TextWrangler, smultron, vim, textmate, notepad++	any	Various. We recommend notepad++ for PC, TextWrangler for mac	You should be very comfortable in finding, opening, saving, and closing text files.
Protégé 4.2 beta 278	all	http://protege.stanford.edu/download/download.html	
IHMC CmapTools v5.04.02	all	http://cmap.ihmc.us/download/	

Phenex 1.2 (most recent)	all	http://phenoscape.org/wiki/Phenex	
Graphviz	all	http://www.graphviz.org/	
Python	all	http://www.python.org/	DO NOT INSTALL YET. To be installed if needed for an advanced component.
	mac	Very likely built in, try "python --version" on the command line.	
	linux	Very likely built in	
Sourceforge Accounts Request Tracking	all	https://sourceforge.net/user/registration	
Gmail Account	all	http://www.gmail.com	
OBO Release Manager	all	http://code.google.com/p/owltools/downloads/list	NOTE: this will be required on day 5 of the course and we will provide complete installation instructions at that time.

Monday, July 30

9am-9:20am **NESCent Introduction**

9:20am-10am **Introductions--who are we? Goals of the course.**

Today is the first day of the anatomy ontology course. The course includes an introduction to ontology principles, including basic logical reasoning, version control, use of ontology editors, and community resources, standards and activities. The course has twelve students in attendance (plus 3 remote participants), ranging from experts in experts in paleobiology and the fin to limb transition, to comparative vertebrate embryogenesis, cartilaginous fishes, weevils, gastropod and cephalopod molluscs, arthropod circulatory systems, taxonomic nomenclature, plant disease, electric fish morphology, mammalian feeding muscle systems, food web representation, and anatomy of model organism such as zebrafish and mouse.

Gaurav: interested in taxonomic names

Nizar: vertebrate anatomy and taxonomy ontologies, extinct reptiles

Eric: comparative genomics and development in marine invertebrates

Laura: botany, plant ontology

Emmanuel: fish evolution, taxonomy

Rambert: PCO, how are others doing ontologies?

Alex: fin to limb transitions, dinosaurs

Christian: circulatory system in arthropods

John: linking atlases

Andrew: taxonomy, weevils

Robert: database of mammalian feeding, oral apparatus muscles

Alexis: evo devo, cartilaginous fishes

Brief introductions by instructors.

Question about what one of the ontology acronyms stand for.

MY: There are a lot of acronyms!

MH: Participants are encouraged to ask questions.

Starting with lots of lecturing to broadly cover the basic principles of ontology logic and anatomical ontology development.

10am-12pm **Overview. The lifespan of an anatomy ontology.**

Slides:	material_for_course/mon/ontologyLifespan.pptx
---------	---

Glossary assignment: Everyone will collaborate on this during the week. Each participant should add a term, a definition, and a comment each day.

Class presentation assignment: 5 minute presentation Friday afternoon. Theme: AO resources or considerations. Compelling problem.

This talk (Matt Y.):

- Workshop flow and philosophy
- Anatomy ontology history and hype
- Lifespan of an anatomy ontology

Mental landscape slide: MH: There is lots of arguing between the groups. People will poke at you to get the answer. Don't take it personally!

Limited concepts: MH: FMA has 80000, CARO has 50; in reality there is quite a variety of size.

Robert: What is meant by "concept"? MH will touch on this more later. This can be put into the glossary.

CLASS EXERCISE: Start adding terms to the glossary (<http://bit.ly/N8Hhc0>).

MH: Many factors contribute to instability in a defined class/concept. Compromises. Anatomical structures change through development and over time. Ontologies themselves evolve over time; they have to have instability or they aren't doing their job.

Nizar: Questions will arise only after ontologies are taking shape.

MH: Zebrafish gene expression: community requirement for structured knowledge, and to answer queries.

RD: With ontology of oral structures, definitions are so generic in order to be applicable across mammals, it is uncomfortable for anatomists.

MH: Ontologies are a tool for retrieving data, so logical, generic definitions are necessary.

MY: And interfaces are still evolving.

1pm-2pm An introduction to (anatomy) ontologies

Slides:	material_for_course/mon/anatomyOntologyReview.pptx
---------	--

Setting the stage (topics) (Melissa H):

- What is an ontology?
- What is an anatomy ontology?
- Anatomy of an anatomy ontology
- Upper ontologies
- How are anatomy ontologies used?

Key take home message for this course: define a concept based on its relationship to other things.

Alex: Is one child enough?

MH: When making subclasses, it is important to identify differentiating characteristics from siblings (will talk more about this later).

Christian: Can't taxonomies do the same thing?

Properties also are referred to as “relations” or “relationships”.

Andrew: Can a class be an instance of a class? e.g. “Alaska” (slide example) could be a class of cities as well as an instance of a state?

Tricky concept.

“Representing development” slide: Alexis: In the EMAP, is e.g. TS13 notochord related to TS14 notochord?

Good question, Terry says both representations exist.

MH: Different design choices depending on what the requirements are.

Eric: Seems like you lose a lot the more you include (e.g. when you are throwing in zebrafish with vertebrates or Uberon and abstracting zebrafish)?

MH: Uberon strategy a little different from GO strategy. Uberon is extremely general, but it is not really for data annotation like the species-specific ontologies, it is meant for reasoning.

2pm-4pm SVN and version tracking

Tutorial:	material_for_course/mon/SVNTutorial.docx
-----------	--

Participants learned the basics of SVN (<http://subversion.tigris.org/>), an important tool used throughout the anatomical ontology community for managing the ontology-related metadata. As part of exercise the participants collaboratively built a “pass-it-on-story”:

It was a dark and stormy nite.

And I forgot an umbrella.

Melissa's cat Peanut is sleeping.

The rain falling on your face...

The roar of the tyrannosaur woke the cat up and it jumped off the sofa.

The cat scratched my face on the way off the couch, it hurt me.

Your mother.

That was a strange dream, must stop eating so late at night!

4pm-5pm Ontological communities and shared principles, trackers, lists, ontology metadata, writing good requests

Slides:	material_for_course/mon/ontologyCommunities.pptx
---------	--

Topics (Melissa H):

- OBO Foundry & Principles
- Bioportal
- OLS
- Ontobee
- QuickGO
- Best practices
- Ontology editors
- Obtaining domain expertise
- Trackers

- Metadata
- Obsoleting classes

Re tracker request example:

Nizar: Actual references more desirable than list of people on a tracker request.

5pm-6pm Modelling, visualizing, and planning your ontology

Exercise:	material_for_course/mon/domainModeling.docx
Cmaps:	group_answers

CLASS EXERCISE: Participants learned to develop design documents for ontologies using CMAP (<http://cmap.ihmc.us>) and parts of the nervous system. It is interesting to see how the classification schemes compare to each other and to existing anatomy ontologies.

HOMEWORK:

1. In a Google spreadsheet, participants were asked to create 10-15 classes from their own domain of expertise.
Link: <http://bit.ly/MRaf4y> (a tab was created for each participant pair).
2. For each class, they were to provide the name, definition, and the differentia that distinguish it from the other subtypes of the same genus. Are there any additional properties that they would want to use?
3. Create Cmaps of classes and properties.

Tuesday, July 31

Continuation of SVN

We illustrated how repository conflicts can be resolved, and files can be moved/deleted and reverted using SVN.

MY: Resolving conflicts in SVN using the class story file. Visualizing differences with graphical tools.

MH: You should never commit a change to an ontology file without looking at the diff.

Participants' Cmaps from Monday exercise

A review of the previous day's ontology design documents that were developed in CMAP. This led to some interesting give-and-take about some specific anatomical problems (nervous system), and ultimately helped to illustrate the point that use-cases and a-priori planning are critical to the developmental goals of an anatomical ontology.

John and Andrew: Optic nerve is a part of CNS, but also usually considered as a cranial nerve which is part of PNS.

MH: You can have two parts of an optic nerve, which are part of CNS and PNS respectively.

MY: Initial stages of planning an ontology: Ask, what is your use case?

MH: You will have more validity in your ontology not based on your domain expertise but because of your requirements.

You can have different kinds of definitions. Text definitions, or even UI specific definitions (depends on your application). Definitions may seem awkward, but sometimes you actually can have different parts of a structure (subclass, make more subclasses).

You can look at requirements, but you can also check what others have done, e.g. look at zebrafish, Xenopus, Uberon. If in Uberon, it may not be right but you know it is present in more than one ontology.

MY: In OWL/Protege, Superclass = SubClassOf

Alexis: Where do you draw the line on how many connections to make between classes?

MH: Look at your requirements!

MY: If you know there's something cool, there is information out there, then you represent it in the ontology.

MH: Which pieces are critical to getting the data you want?

MY: When starting your ontology, focus on just a few properties (not the entire list e.g. that you see in Uberon in Protege).

Usually start with is_a relationships as the anchor for your ontology, then consider others such as part_of. Protege requires that everything is classified.

10am-11am Anatomy of an OBO file, Intro to OBOEdit

Slides:	material_for_course/tues/ontologyOBOFileAnatomy.pptx
---------	---

The primary topic of the day was an introduction to the OBO world.

Presentation by Matt Y:

OBO file format was designed to be easily readable by both machines and humans.

[Typedef] is code for property, or relationship, relation, etc.

In the Tree Editor, read backward to get the relationship (Talking Parrot is_side_kick_of Mad Scientist, not the other way around).

Tree Editor mixes together all the relationships, Protege by default only shows subclass relationships. All properties/relationships are indented. Some terms show up repeatedly because of this.

Key bits:

- tags
- stanzas
- dbxrefs

Dbxref = definition reference

Xref = equivalence to a class in another ontology or source

A basic file:

- header
- term
- typedefs

11am-1pm **Navigating an ontology in OBO Edit, basic ontology creation in OBO edit**

Tutorial:	material_for_course/tues/OBOEditTutorial.docx
Answers:	material_for_course/tues/OBOEdit_tutorial_answers.docx

CLASS EXERCISE – An OBO tutorial focused on search, queries, and navigation. This allowed course participants to have an introduction to some basic reasoning, as well as to learn some of the visualization tools available in OBOEdit that are not available in Protege.

We were using the latest version of OBOEdit, and were thus acting as beta testers. Unfortunately this resulted in an unstable platform for a few of our Windows-using participants.

2pm-3pm **Using anatomy ontologies for annotation. Linking in genetic data.**

Slides:	material_for_course/tues/PhenotypeAnnotation-anatomycourse.pdf
---------	---

We moved to a lecture and demonstration lead by Jim Balhoff on phenotype annotation using the “EQ” syntax, and post-composition. Jim demonstrated the basics of Phenex configuration and use and also some examples derived from the HAO’s approach to instance and phenotype modelling with respect to semantic phenotypes.

MH: There are a lot of curator inconsistencies. Having definitions of terms is not enough. Need high quality guidelines that tell curators how to apply the annotations. You can use logic of ontologies, however, to reconcile and converge on an annotation.

Rambert: Clarify pre vs post composition?

JB: It comes down to deciding whether or not to include particular classes in your ontology if they can be defined in expressions. e.g., you could add “head bristle” to the ontology, or you can say “any seta that is part_of the head”; setae can be part of many things.

MH: Using logic to make your definitions is what you really want for data retrieval.

Robert: Clarify classes vs instances vs annotations?

MH: Will be addressed throughout course. Annotations can be applied to a class or an instance. ZFIN curated data example: They are really curating class, no identifiers given to the specimens that are being described.

Protege example: Instances can be intermingled with classes in the ontology.

Queries are based on classes, logic is based on instances.

3pm-5pm

Curating evolutionary phenotype data using Phenex

CLASS EXERCISE

What characters were troublesome in this process?

- Limitations in terminology -- lots of detail you can't capture, but it's ok not to capture every detail. MY: Phenex is suited to fairly general legacy data. Perhaps it is not needed for more detailed, specific data.
- Pre-composition. Ask, is it more work to create classes in the ontology, or for annotators to keep composing terms? Your decision, the reasoners don't care.

ZFIN: discussion of curation of gene expression. Effort to persuade and require investigators and journals to organize, annotate data, etc. Incentive to do this in a more open world. This is going to change how curation is done in the future.

Continuation of Cmaps, stubbing ontologies

CLASS EXERCISE

Wednesday, Aug. 1

Instructors plan to meet for ~5 minutes with each participant to discuss practical outcomes for each student in the course.

Comment from MY and MH: OBO Edit still has use and currency -- ability to change colors, visualization, rendering graphics for publication. Different ways of viewing the ontology for quality assurance. We want data to be agile.

Today is completely devoted to an introduction to the Web Ontology Language, OWL. After some morning theory, participants will dig into a 41 page tutorial (as well as 5 additional exercises) on the use of Protégé.

9am-10am

Intro to OWL

Slides:	material_for_course/weds/OWLintro.pptx
---------	--

Topics (Carlo T):

- What is OWL?
- IRI, URI, URLs and URNs
- OWL document, syntaxes
- RDF/XML
- Ontology document
- Namespaces
- Ontology header
- Entities: individuals, classes, properties

- Disjointness
- Open World assumption
- OWL properties: object, datatype, annotation
- Protégé
- Reminder: when you manually edit an OWL file...you go to OWL Hell!

RDF: Subject → Predicate → Object

Annotations, properties...everything is a triple.

Even when making a synonym, you are saying: *A is a synonym of B*

Eric: What is the XML side of RDF/XML?

CT: It is a set of structures, a markup language (HTML is a subset of XML).

e.g., a cmap can be interpreted as an RDF graph, but when you write it down, XML gives you a syntax and grammar.

OBO format: single, easy to read file format; this is not going to happen in OWL.

Point of presenting namespaces is not to memorize them, but to demystify them if you need to look in a file later and troubleshoot it.

If you want to see the triples, it is difficult to do so in the RDF/XML code.

Eric: Are all of these entities residing on the web (since the IRI is an http address)?

CT: Not necessarily. But they should be resolvable on the web once you release the ontology to the public.

They then become an attributable entity.

That is the power of the semantic web.

purl = persistent URL. We will going through setting this up on Friday.

In OBO everything is assumed to be disjoint. In OWL you have to make that declaration.

Always define domain and range when you can. "Close the world!"

In moving bio-ontology developers from the OBO to OWL world, two issues:

1. People are still thinking of individuals as classes.
2. There is a lot of work to do in closing the world with domain and range.

<u>Domain</u>		<u>Range</u>	
A's	my_relation	B's	
Then:			
some C	my_relation	some B	NO
some A	my_relation	some B	YES

10am-12pm The big Protégé tutorial

Tutorial:	material_for_course/weds/tutorial/protege_tutorial.doc
-----------	--

CLASS EXERCISE

Reminder to everyone to quit and restart Protege whenever they need to open a new file.

1pm-5pm **Continue with prior tutorial based on time. Explore anatomy ontology exercises.**

CLASS EXERCISES

5pm-5:45pm

Slides:	material_for_course/weds/musclesBones.pptx
---------	--

MH and CT (slides and example in Protege): Understanding the difference between declaring a property and class restriction.

Will do another exercise integrating instances.

Thursday, Aug. 2

It is recommended that everyone go back and do the Protege “Pizza” Tutorial after the course, will help jell some of the concepts.

<http://bit.ly/Q5NHj>

Today, with OWL basics in hand, we move on to a tutorial about importing other ontologies: where and how to get them, and how to tell Protégé what to do with them. It is not always convenient to import a whole other ontology, so we will learn the principles of “MIREOT”, or the minimum information required to reference an external ontology term. This requires the complex navigation of the Ontofox tool to obtain bits and pieces of external ontologies. We then learn how to put them back together again (in Protégé).

9am-10am **OWL: imports and external ontologies**

Slides:	material_for_course/thur/imports_mireot_ontofox.pptx
Tutorial:	material_for_course/thur/tutorials/imports_tutorial.doc

Topics (Carlo T):

- OWL imports
- Imports in Protégé
 - Ontology libraries and catalog files
 - MIREOT
- OntoFox
 - Examples

MH: Why might you like to have local files of the ontologies you are working on? They are dynamic and changing, so you might want a “snapshot” of them on your desktop. If they are changed by another user, you can decide how to resolve.

CLASS EXERCISE: Imports tutorial.

10am-10:30am Intro to Ontobee and imports

Slides:	material_for_course/thurs/Ontobee.pptx
---------	--

Looking up ontology terms with Ontobee (Carlo T):

- What is Ontobee?
- Ontobee vs. Bioportal
- Hands-on: <http://www.ontobee.org/index.php>

10:30am-12pm Ontology alignment, interoperability, reuse, building compositional anatomical entities from different ontologies

Slides:	material_for_course/thurs/OntologyInteroperability.pptx
---------	---

Topics (Melissa H):

- How to synchronize anatomy ontologies
- Three approaches:
 - mapping
 - Xref strategy
 - imports/MIREOT
- Proposed model
- Modularizing ontologies – what we need
- Idealized protocol for new AOs

You can create a placeholder term in your own namespace if you need something to reference, even if it is outside of your domain, then request the term in the appropriate ontology (e.g. a biological process in GO). Later you can import that term and obsolete the one in your namespace.

Use of “sensu” when associating labels with things. Separate classes may have been given the same labels. You can add annotations to each class to make the old label a synonym.

We need to encourage the community to make annotations of annotations (such as synonyms) to identify where they come from. This hasn't really happened in the OBO world.

Presentation of test_composition.owl (svn/trunk/materials_for_course/thurs/using_multiple_ontologies) in Protege. Note all of the obsoleted classes. HAO classes that were originally created and xrefed to other ontologies have now been replaced by directly imported classes.

It is a good practice to add comments when obsoleting classes.

In Protege, put into practice an example from the slides (slide 18):

- Run DL query: (seta and bearer_of some white) and (part_of some head)
- Add to ontology and run reasoner

- Inspect what the new class is a subclass of

1pm-2pm **MIREOTing, Ontofox**

Tutorial:	material_for_course/thurs/tutorials/ontofox_tutorial.doc
-----------	--

CLASS EXERCISE: Ontofox tutorial.

CT: Note that you can upload your saved Ontofox input file in order to repopulate the web interface (section 2).

It is often easier to modify the text file because of the difficulty in using the Ontofox interface (it's terrible!).

3pm-4pm **Protégé plugins. Image annotation plugin. Phenotype annotations in Protege sensu HAO. Adding images to your ontology. Annotating classes.**

Slides:	material_for_course/thur/protege_images.pdf
Tutorial:	material_for_course/thur/tutorials/individuals_tutorial.docx

We saw a classification of the Balhoff family tree in Protege. We learned to classify instances and use them for testing our anonymous classes. We learned to install Protégé plugins and link images within our ontology, including pictures of Jim's baby, hymenoptera heads and blue jeeps.

Image depiction plugin URL: <https://github.com/balhoff/image-depictions-view>

4pm-5:30pm **Spatial exercise**

CLASS EXERCISE

/svn/trunk/materials_for_course/thurs/using_multiple_ontologies/bspo.owl

We broke into groups to brainstorm what would be needed to represent spatial relations and classes for the postcomposition of anatomy classes. This was done without peeking at the Spatial Ontology, in order to gain unbiased requirements for this ontology and get the participants thinking about spatial relationships.

MH: Spatial classes were built to be used with anatomical ontologies.

MY: The ontology is flat, they are all siblings.

(The Usage tab in Protege is very important. It shows you all the things in your ontology that reference the Thing you are currently on.)

There are axes and sides, but also consider attachments and connections.

Compartments, regions, and areas -- well-defined vs. not well-defined boundaries.

Boundaries are often a site of gene expression.

There are bona fide boundaries and fiat boundaries.

Eric: How to represent amorphous, scattered, disorganized (though there may be underlying developmental organization)?

Alex: Boundaries: consider strong bounded, weak bounded.

Spatial classes are very modular building blocks.

Friday, Aug. 3

Follow up to Thurs. Spatial exercise

Where was disjointness used in the Spatial Ontology? How to find out:

Sparql -- Semantic SQL. Syntax is a lot like SQL (SELECT, WHERE, etc) and lets you query the Semantic Web.

It is in a tab in Protege, but not specific to it. When you have data in a triple store database, this is the way you would query it.

(There are no disjoint classes in Spatial because it came from OBO.)

9am-10am Discussion of group projects

Files:	user/participant_name
--------	-----------------------

We started the the last day of our course with presentation and discussion of group ontology exercises and issues facing the different domains, such as nomenclature and a nice example of linking taxon-specific anatomies with mineralized properties of tissues. We discussed an evolutionary time-analogy to developmental time for paleontology, which resulted in the question "Is the big bang the first temporal boundary?"

Eric and Andrew:

Arthropods and mollusks.

Was there anything we can imagine that would be under the same class?

A number of differentiating characteristics for the subclasses of calcium carbonate biomineralized structure.

Natural inclination is to make the ontology spatial. But when doing it comparatively, we did not approach it spatially at all. Looking at properties instead.

Narrowed to one structural axis -- keeps it focused.

Eric: Would rather everything hang off of the partonomy.

MY: Don't think hierarchically!

Andrew: It's more of a networking system than hierarchical.

Nizar and Alex:

A geology ontology, with a time perspective.

Time period ← Cretaceous ← various times ← formations that have fossils

Are formations classes or instances?

Paleoenvironments: formations link to these.

Using data from phenoscape, e.g. look for certain morphologies at certain times or environments.

You can also add information about paleo-latitude. Can give you interesting info about where particular morphologies evolved.

MY: Remember differentia, e.g. between Chronological unit, Period. These aren't words, they are concepts.

Christian:

When I came here, I wanted to do an AO that was as simple as possible.

Defining more and more relationships between classes. Eric argued that this makes it more rigid?

MH: Argues that the opposite is true, it makes it less rigid, you're not locked in a partonomy. They are a test for the definitions.

MY: With HAO, by forcing ourselves to write definitions out that followed a genus-differentia pattern, those extra properties jumped out.

10am-12pm

Releasing ontologies in multiple formats using the Oort tool

Slides:	material_for_course/fri/releasing_ontologies.pptx
Tutorial:	material_for_course/fri/tutorial/tutorial_ontology_release.doc

We discussed what it takes to release an ontology in different flavors, and decided that a flavor is a slice of a version. These are end goals to keep in mind as we go about our work in developing anatomy ontologies. It is important to know what activities are necessary to make your ontologies public.

Topics (Carlo T):

- What does it mean to release an ontology?
- Suggestions
- Recommendations
- Tools
- Example

PURLs and IRIs. Do NOT EVER publish a URL that is not a PURL. You can always update where you point them. And remember, you are a reviewer too! What do you see in the publications you are reviewing? Tell authors that they should uniquely reference ontological entities (among other research resources), that these should be referenced using PURLs, and that they should be attributed.

What to do if you accidentally release your ontology with a reasoning error? Answer -- change your email address! No, fix the error in the release and apologize :-)

Lately we have stressed pushing requirements to the Protege community.

MH: Remember, bio-ontologies are just a small part of the Protege world, others are using it in very different ways.

CLASS EXERCISE: Oort tutorial.

1pm-2pm Homology 1, issues and navigating through ontologies that formalize homology hypotheses. Homology 2, modern sounds in homology reasoning.

Slides:	material_for_course/fri/homology.pptx
---------	---------------------------------------

We addressed the beforehand “taboo” topic of homology. The presentation of how the community is currently representing homology hypotheses external to the anatomy ontologies seemed to go over without too much controversy.

Topics (Matt Y):

- Design considerations
- Consequences
- Stewardship

Homology is about details, which we circumspect, score, and robustly define.

Is it dangerous to mix function and structure?

MH: You can define your AO however you want, e.g. by function.

JB: OWL lets you define sets of things. But that’s all they are, sets of things. Try to focus on what the class definitions are.

We’re using the anatomy ontology to test the logical completeness of our anatomical statements, not our homology statements.

Semantic phenotypes are very diverse.

There is a lot of diversity in the descriptions. Speaks to the need for ontologies to have robust annotation guidelines for how to apply them.

Homology will derail discussions!

MH: We have been working on homology relations for 5 years to define it logically. It is extremely difficult.

Talk by Jim B:

http://phenoscape.org/wiki/Reasoning_over_homology_statements

MH: You have to define things in terms of their descent.

2pm-3:30pm Student presentations

We concluded the week with 5 minute lightning talks of students discussing some problem, perspective, or plan. These turned out to be a great highlight of the week, and it was suggested that they should be emphasized in future years of the course.

Gaurav: Interested in species and taxonomies and their representation in ontologies. Species change and species boundaries change, so how can ontologies deal with that? Museum specimen example. What happens when someone comes along and describes a new species and a species gets split?

Eric: Does this need an ontology, or is it a mapping problem?

MH: There is a data provenance problem.

Nizar: Homology meeting in Chicago recently, agreed there are different types of homology and homology statements. Difficulty in incorporating functional terms, e.g. for flight in birds. Re the

Vertebrate Anatomy Ontology, will be talking about this project at an upcoming meeting. How do you get people in the field to get interested in building and improving bio-ontologies? Is that going to happen?

MH: At ZFIN, small task forces helped developed parts of the AO, worked much better than large-scale effort. What they wanted was attribution.

Eric: Key is GUIs, making user interface to Protege more friendly.

CT: There are developers and domain experts. How to bridge them? Protege is not the tool for that. Spreadsheets, MX, etc can help.

Robert: Need interesting use cases.

Laura: Has a list of plant diseases that are categorized. Goal is to make an ontology that describes them by common name and what plants they affect. Issue is that making a class for each disease+plant name combination seems redundant and difficult. Same disease may be doing different things to different plants -- how do you represent this?

CT: you could use an "affects" property.

Alex: Adding to what Nizar talked about. Involved in Phenoscape. Huge amount of work needed to just get started with necessary analyses. Need use cases, so you have something to show people and solicit their help. Perhaps start with the "biggest and baddest" publications now to get people on board, add the details later. This will also inform priorities wrt to improving tools like Phenex.

Emmanuel: Has a data matrix, with 50 species and 107 characters, wants to put them into an ontology. Started by building a cmap of a skeleton. Got terms translated into OBO and then converted into OWL format. Now wants to figure out how to formalize characters and character descriptions in an ontology. Challenges are homology and character state descriptions, loss of freedom because of formalization, and "Darwin's pebbles" problem.

CT: Ask, will ontologies give you something other tools won't give you? Scope out your requirements.

Rambert: Working with a graduate student who is describing interactions between species in a system. How to incorporate mathematical elements into an ontology? Organisms she is dealing with are deer, bees, wasps, and plants. Has looked at BFO, CARO, but no AO for wasps... wrt encoding all of these interactions, there are some ecologically related words in GO, but wants to talk to GO to bring an ecologist's perspective.

Christian: Works on the circulatory system of arthropods. Coming from a morphologist's standpoint. There is a challenge to morphology's relevance from all the genomic data now available. Wants to stay on the descriptive level and use ontologies to infer evolutionary hypotheses. Showed beautiful slides of circulatory system parts.

MH: We want to see the ontology with those images!

John: Working on human embryo project: serial sections of the 23 human embryonic stages and annotation of images with structures. Wants to be able to search the images with those labels. Wants to tie together the human atlas and the mouse atlas. Problem is classes being repeated for each stage as in the mouse anatomy.

MH: There is a way to collapse this down. Can coordinate with Terry.

Andrew: Coleoptera AO. Formalizing species descriptions and wants to make them interoperable with computers. Ontologies could be a powerful tool. But some problems: taxonomic literature not computer friendly, dealing with natural language, nomenclatural variety,

limitations of tools. CoIAO is in MX. Can tie in phylogenetic analyses. Needing to learn about object relationships (is new to ontologies).

Robert: FEED project to create online repository of physiological studies of feeding in mammals. Huge amount of synonyms in literature for concepts. Thinking of different parts of the muscles (e.g. masseter), that are consistent across the mammals. Looked at distribution of characters across mammals. With Paula and Hilmar at NESCent, came up with relationships and generic definitions. Has gone from Spreadsheet → Phenote → Protege.

3:30pm-4pm **Course evaluation**

Reception, dinner and social.